1360 Redwood Way, Suite C
Petaluma, CA 94954-1169
707/665-9900
FAX 707/665-9800
www.sonomatech.com

ESTIMATION OF BACKGROUND CONCENTRATIONS FOR NATA 2002

DRAFT FINAL REPORT

STI-906206.05-3291-DFR

By:
Mike McCarthy
Juli Rubin
Bryan Penfold
Hilary Hafner
Sonoma Technology, Inc.
Petaluma, CA

Prepared for:
U.S. Environmental Protection Agency
Raleigh, NC

January 2008

ESTIMATION OF BACKGROUND CONCENTRATIONS FOR NATA 2002

DRAFT FINAL REPORT
STI-906206.05-3291-DFR

By:
Mike McCarthy
Juli Rubin
Bryan Penfold
Hilary Hafner
Sonoma Technology, Inc.
1360 Redwood Way, Suite C
Petaluma, CA 94954-1169

Prepared for:
James Hemby
U.S. Environmental Protection Agency
USEPA Mailroom, Mail Stop C304-04
Research Triangle Park, NC 27711

January 4, 2008

This PDF document contains blank pages to accommodate two-sided printing.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

OVERVIEW

The National Air Toxics Assessment (NATA) is an ongoing effort by the U.S. Environmental Protection Agency (EPA) to identify and prioritize air toxics, emission source types, and locations that are of greatest potential concern in terms of contributing to population risk. The NATA results are used by EPA to help communities design their own local-scale assessments, to set priorities for improving data in emissions inventories, and to help direct priorities for expanding and improving the network of air toxics monitoring. The most recent national-scale assessment focused on the 1999 air toxics emissions inventory, which was completed in December 2005 (http://www.epa.gov/ttn/atw/nata1999/). One of EPA's goals for the NATA program is to continually enhance and improve the modeling and assessment tools. For the most recent assessment, NATA 2002, EPA seeks to develop refined estimates of background concentrations of hazardous air pollutants (HAPs, also called air toxics). This document describes the methodologies used to develop estimates of HAP background concentrations for the NATA 2002.

INTRODUCTION

NATA 2002 uses the Assessment System for Population Exposure Nationwide (ASPEN) model to estimate concentrations of air toxics by census tract in the United States. The modeling methodologies in the national-scale assessment estimate long-term outdoor concentrations of air toxics attributable to 2002 anthropogenic, or human-generated, emissions. However, the ASPEN model does not estimate outdoor concentrations of air pollutants attributable to long-range pollutant transport, unidentified emission sources, and natural emission sources. These "background" contributions can be significant for some air toxics and should be accounted for to accurately model concentrations. For NATA, background concentration estimates are those concentrations reflecting transported contributions from farther than 50 km, unidentified emissions sources, and natural emissions sources.

Two previous NATA exercises were performed for the model years 1996 and 1999. These iterations used background concentrations developed specifically for NATA. In the 1996 exercise, background concentrations were gathered in a literature search performed as part of the Cumulative Exposure Project (CEP). The CEP literature review was originally performed to acquire background concentrations for 1990. The result of the literature search was a single remote background value representing 12 air toxics.

In the 1999 exercise, two approaches were used to estimate background concentrations. The primary approach estimated background concentrations using measurements from ambient monitors. Estimates from individual locations were extrapolated to counties without measurements based on a population regression. When ambient measurements were not available from the ambient monitoring network, background concentrations from the CEP were used.

Table 1 lists the pollutants for which background concentrations were estimated for NATA 1996 and 1999. This table is adapted from the NATA 1999 web site.[1]

---

[1] http://www.epa.gov/ttn/atw/nata1999/99pdfs/backgroundtable.pdf

Table 1. Pollutants for which background concentrations were estimated in the NATA 1996 and 1999 modeling exercises. Pollutant names in italics were assigned values from the CEP literature search and were not spatially variable.

NATA 1996	NATA 1999	
Benzene	Benzene	1,3-Butadiene
Carbon tetrachloride	Carbon tetrachloride	*Bis(2-ethylhexyl)phthalate*
Chloroform	Chloroform	*Bromoform*
Dichloromethane	Dichloromethane	*Carbon disulfide*
Ethylene dibromide	Ethylene dibromide	*Chlordane*
Ethylene dichloride	Ethylene dichloride	*Hexachlorobutadiene*
Formaldehyde	Formaldehyde	*Hexachloroethane*
Mercury	*Mercury*	*Lindane*
Polychlorinated biphenyls	*Polychlorinated biphenyls*	*Methyl bromide*
Tetrachloroethylene	Tetrachloroethylene	*Methyl chloride*
Trichloroethylene	Trichloroethylene	*Methyl chloroform*
Hexachlorobenzene	Acetaldehyde	*Phosgene*
	1,1,2,2-Tetrachloroethane	Vinyl chloride
	1,2-Dichloropropane	*Xylenes*

TECHNICAL APPROACH

Two methods were used to develop estimates of background air toxics concentrations for the NATA 2002. The first method relies on ambient air toxics measurements (ambient-based method) and the second method relies on HAPs emission inventory data (emissions-based method). The ambient-based method is preferred because the background estimates are based on measured air toxics concentrations throughout the United States. However, reliable ambient measurements are not always available for every pollutant of interest. Therefore, an emissions-based method was developed to handle those pollutants with inadequate ambient measurements. In addition, a few pollutants were assigned uniform spatial concentrations based on their long lifetimes and well-characterized concentrations. These pollutants are carbon tetrachloride, methyl chloride, methyl bromide, and methyl chloroform. All are routinely measured at remote sites and have well-mixed concentrations in the Northern Hemisphere[2] (Montzka et al., 2000; Montzka et al., 1999; Prinn et al., 2000). The values for these pollutants are presented as standard estimates below (Ambient-based Methodology section).

Method Selection

Prior to developing background concentration estimates, an assessment of ambient air toxics data availability and quality was performed to identify the pollutants that were candidates

---

[2] ftp://ftp.cmdl.noaa.gov/hats/

for the ambient-based method and those that would require the emissions-based method. Two criteria were used to determine if adequate data were available for a pollutant for the ambient-based method:

1. For a given pollutant, at least 100 ambient measurement locations were required for adequate spatial representativeness. In general, 100 monitoring sites resulted in a reasonable spatial distribution of monitoring locations for most pollutants.

2. For a given pollutant, at least 85% of the ambient measurements had to be above the method detection limit (MDL). The MDL is a measurement process characteristic that establishes the level at which a reported measurement is considered to be statistically significantly greater than zero. Concentrations reported at or below the MDL have a high relative uncertainty. When a high percentage (>85%) of measurements for a given pollutant are below the MDL, the average pollutant concentration has a high relative uncertainty and is considered to be poorly characterized. Ambient measurements that are consistently reported below MDL are unlikely to provide useful quantitative information for estimating spatial variability in background concentrations.

The two criteria listed above were applied to the ambient measurements for 30 air toxics of interest. The ambient-based methodology was used to estimate background concentrations for those species with ambient measurements that met both criteria listed above.

The emissions-based method was applied to species for which ambient data failed to meet the ambient-based method criteria. One exception was chromium, which was characterized using both methods. Chromium has a specific oxidation state (hexavalent chromium) that is toxic, but most ambient measurements do not distinguish its oxidation state and simply measure total chromium. The ambient measurements reflect this total chromium value, while the emissions-based method specifically represents chromium VI. **Table 2** lists the air toxic pollutants for which background estimates were developed and the methodology used for each.

Data Availability

Ambient air toxics data were acquired for 2002 through 2005 from EPA's Air Quality System (AQS). These data were supplemented with measurements from the Interagency Monitoring of Protected Visual Environments (IMPROVE) and the SouthEastern Aerosol Research Characterization Study experiment (SEARCH). Data from AQS were downloaded in July 2007, while the supplemental data were previously acquired in August 2006.

Air toxics measurements are primarily collected as 24-hr duration samples. These samples are most often collected at 1-in-3, 1-in-6, or 1-in-12-day frequencies. Any samples collected with less than 24-hr duration (e.g., 1-hr or 3-hr samples) were aggregated into 24-hr averages if at least 75% of the day was measured. For example, at least 18 1-hr samples were required for aggregation to a 24-hr average. This ensured reasonable diurnal pollutant representation.

The 2002 National Emissions Inventory (NEI) was acquired and used for the emissions-based method. Total emissions by county and pollutant are reported in the NEI.

Table 2. Methods used to estimate background concentrations for HAPs of interest for NATA 2002.

Ambient-based Method	Emissions-based Method
1,3-Butadiene	Hydrazine
1,4-Dichlorobenzene	Chromium (VI)
Acetaldehyde	Ethylene Dichloride
Arsenic	Naphthalene
Benzene	Propylene Dichloride
Chloroform	Ethylene Oxide
Chromium	Acrylonitrile
Dichloromethane	Cadmium
Formaldehyde	Beryllium
Lead	Ethylene Dibromide
Manganese	Benzidine
Nickel	Quinoline
Tetrachloroethylene	Bis(2-Ethylhexyl)Phthalate
Toluene	1,2-Dibromo-3-Chloropropane
	Trichloroethylene
	1,1,2,2-Tetrachloroethane

Ambient-based Methodology

Background concentration estimates for 14 species were developed using the ambient-based method. This method consists of four general steps:

1. Create annual average concentrations and select monitoring sites with seasonally representative measurements for each of the 14 species.

2. Determine the quality of measurements for each pollutant at each monitoring location and calculate initial background concentrations.

3. Quality assure (QA) and quality control (QC) the results.

4. Apply standard background values to areas lacking ambient measurements.

Create Annual Averages and Select Representative Monitoring Sites

Ambient concentration data from AQS, IMPROVE, and SEARCH were aggregated to annual averages using completeness criteria for the 14 species listed in column 1 of Table 2. To create annual averages that adequately represent the entire year, quarterly calendar averages were created first.

4

Only 24-hr averages, based on both sub-daily data and 24-hr duration measurements, were used to create the quarterly average data. The quarterly averages were computed as follows:

- Calendar quarters were defined as January through March, April through June, July through September, and October through December.

- Quarterly averages were calculated using a 75% data completeness criterion based on the sampling frequency for a given monitoring location. If sampling frequency information was not available, a minimum of six valid daily average values were required per quarter for each monitoring location; this minimum count coincides with 75% completeness criterion applied to a 1-in-12-day sampling frequency.

Annual averages were then calculated by averaging the quarterly averages at a monitoring location. At least three of four quarterly averages were required to adequately represent the seasonal variability in pollutant concentrations.

We ensured that individual measurements represented the seasons of the year for a given location and pollutant by requiring a valid annual average at a monitoring location. This requirement meant that at least three of four seasons are represented with at least six measurements for each. Separately, we required 30 samples at each monitoring location to provide slightly more robust statistics for the percentiles used (i.e., 30 samples will provide a better number than 18 samples when trying to determine a 10th percentile value or the percentage of samples below MDL). The actual background estimation was then performed using 24-hr averages as described in the next section.

Determine the Quality of Measurements for Each Pollutant at Each Monitoring Location and Calculate Initial Background Concentrations

Data for each site and species were assessed to determine the number of samples below the reported MDL. Each monitoring site was then assigned to one of three MDL bins based on the percentage of samples reported below the MDL: (1) <10%, (2) 10-85%, and (3) >85% as shown in **Table 3**. Sites with less than 10% of data below the MDL were assigned to the first bin. For these locations, the 10th percentile concentration reported at that site was used as the background concentration. The 10th percentile concentration corresponds to the cleanest days monitored at a site, which we would expect to be representative of "clean air" background concentrations. Choosing the 10th percentile rather than another small percentile will have a relatively small influence on the final background results. Given the typical log-normal distribution of pollutant concentrations at monitoring locations, the difference between the 5th, 10th, and 20th percentile concentrations are small at the lower end of the distribution.

Sites with 10-85% of data reported below the MDL were assigned to the second bin. When 10-85% of data are reported below the MDL, the 10th percentile concentration is considered unreliable; therefore, an alternate method was required to estimate the 10th percentile concentration for sites assigned to the second bin. For those sites, the fraction of samples reported above the MDL was multiplied by the MDL to estimate the 10th percentile concentration. For example, if 35% of samples were above the MDL, the estimated background

concentration would equal (0.35)*MDL. This approach ensured that the percentage of samples above the MDL for a given site was reflected in the background concentration estimates.

Table 3. MDL bins and corresponding methods used to calculate initial background concentration estimates for each site and pollutant.

Percent of Data Below MDL	Estimation Method
<10%	Use 10th percentile concentration
10-85%	Fraction of samples above MDL * MDL
>85%	0.10 * MDL

Sites with at least 85% of data reported below the MDL were assigned to the third bin. The equation (0.10)*MDL was used to estimate background concentrations for sites assigned to the third bin because data that fall below the MDL do not provide useful information about the distribution of concentrations at a site and are, therefore, treated homogeneously. Of the three MDL bins, the sites assigned to the third bin are considered the least representative for estimating background concentrations.

After the sites and species were assigned to MDL bins, background concentrations were estimated using the methods discussed above. Summary statistics were then calculated for each pollutant including 5th, 20th, 50th, 80th, and 95th percentiles. The summary statistics were used to QA and QC the background estimates and to assign background concentrations to counties without ambient measurements.

Quality Assure and Quality Control the Results

A flowchart illustrating the QA and QC methods used to assess the background concentrations is shown in **Figure 1**. Statistical and visual tests were performed to QA and QC the site-specific background estimates:

- Comparing results from each location to remote background concentrations. Concentrations below the remote background concentrations were replaced with the remote estimate.

- Identifying all sites with background concentrations more than two times greater than the 80th percentile background concentration. These estimates were replaced with the 80th percentile background concentration.

- Flag the highest and lowest 5% of sites for further visual inspection.

- Identifying all sites with background concentration estimates below the MDL when the MDL is more than twice the national average. These data were flagged for further visual inspection.

- Flagging all locations with background concentration estimates above the MDL when the background concentration is above a health benchmark threshold (i.e., 0.5-in-a-million cancer or 0.1 hazard quotient) for further visual inspection.

6
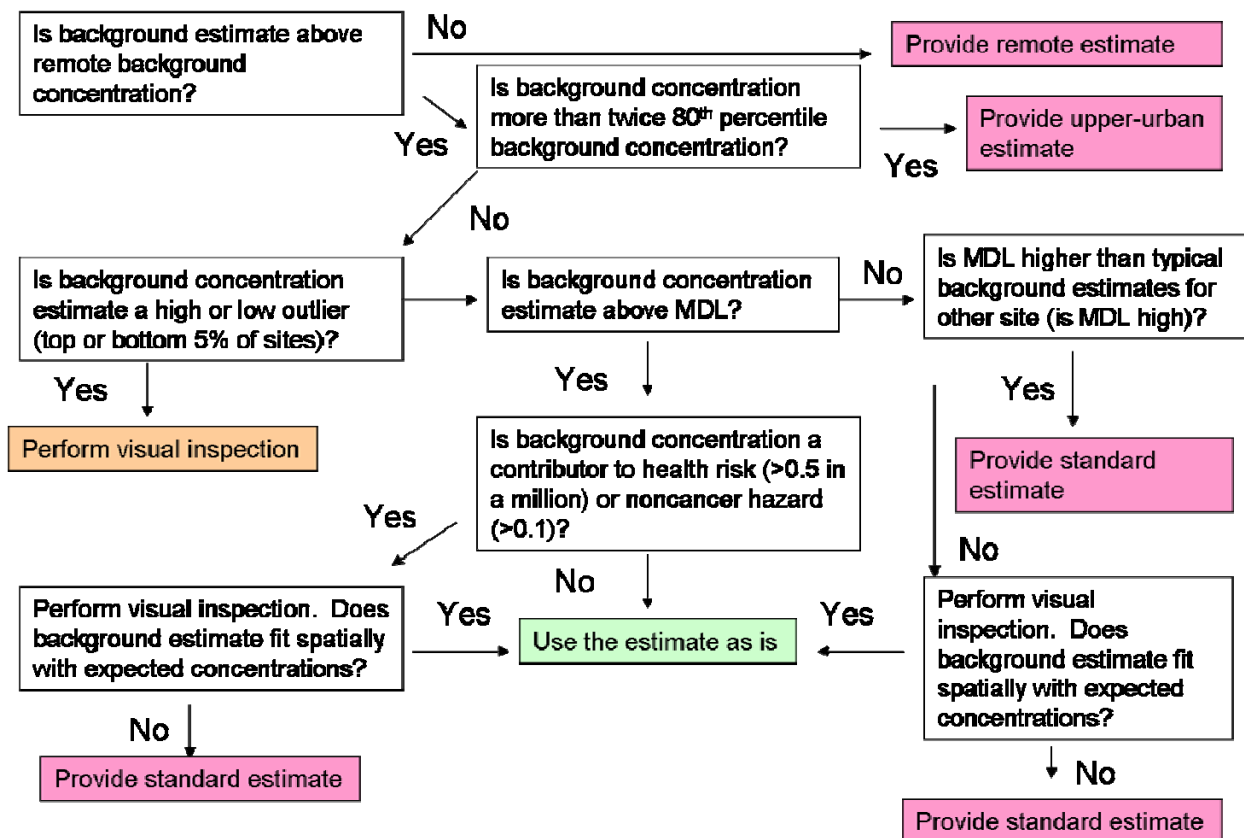
Figure 1. Flow chart illustrating the QA and QC steps performed on site-specific background concentration estimates.

Sites that appeared to be suspect were then further inspected by an analyst using maps of background concentration estimates, average MDL values, and average concentration values for all locations. **Figure 2** is an example of a map used to visually inspect concentrations for benzene. Visual inspections included the following steps:

- Compare background estimates to average pollutant concentrations for a particular location. The background estimates should be between 10-50% of the average concentration for most pollutants. Most sites that had high background concentrations relative to average concentrations were in the >85% below MDL bin.

- Inspect background concentration maps to identify spatial patterns in concentrations. Areas with high or low background concentrations that appeared inconsistent with regional patterns were flagged.

- Inspect individual outliers with high or low background estimates. Locations that appear unrealistic given the spatial patterns of concentrations were flagged.
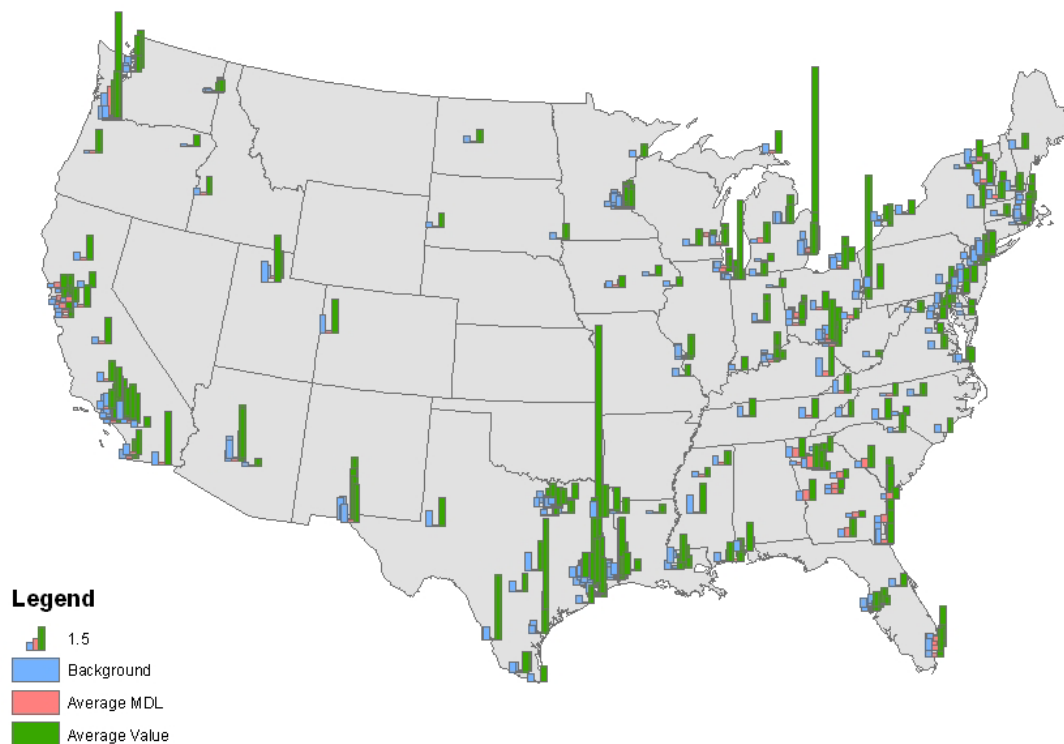
Figure 2. Example visual inspection map for benzene. Background estimates, average MDL values, and average concentrations (all in µg/m³) are displayed for all sites with monitoring data (2002-2005).

After performing QA and QC on the site-specific data for the 14 pollutants, county estimates were generated. Because monitoring sites do not exist in all counties, and/or some counties are quite large and may only have a single monitoring location, the background estimates were applied to the county level. The following rules were used to create the site-to-county assignments:

1. For all counties with only one monitoring location, the estimate for that location was assumed to represent the county.

2. For counties with multiple monitoring locations, the location with the minimum concentration was used. The minimum concentration was assigned as the background to provide an estimate of the lowest county concentrations.

3. Counties with no available ambient measurements were assigned a standard estimate (discussed below).

Apply Standard Background Values to Areas Lacking Ambient Data

Because background concentration estimates are influenced by many factors, some estimates were unlikely to be reliable or representative of the real atmosphere. For example, some sites had very high MDL concentrations that would result in high background estimates using our approach. For these and similar cases for which the reliability of the background

8

estimate is questionable, a "standard" estimate was assigned to the site. In addition, many counties had no available ambient data to estimate background concentrations. For these counties, the standard estimates were assigned.

Standard estimates were defined for four population bins based on the population distribution of U.S. counties and the subset of counties with ambient HAP monitors. The bins and definitions are as follows:

1. Remote—population < 25,000 and rural county

2. Small urban—population between 25,000 and 100,000 and rural, or population < 25,000 and classified as an urban county by NATA 1999

3. Medium urban—population between 100,000 and 1,000,000

4. Large urban—population > 1,000,000

Figure 3 illustrates the population distribution of U.S. counties and the subsets of counties with ambient measurements.
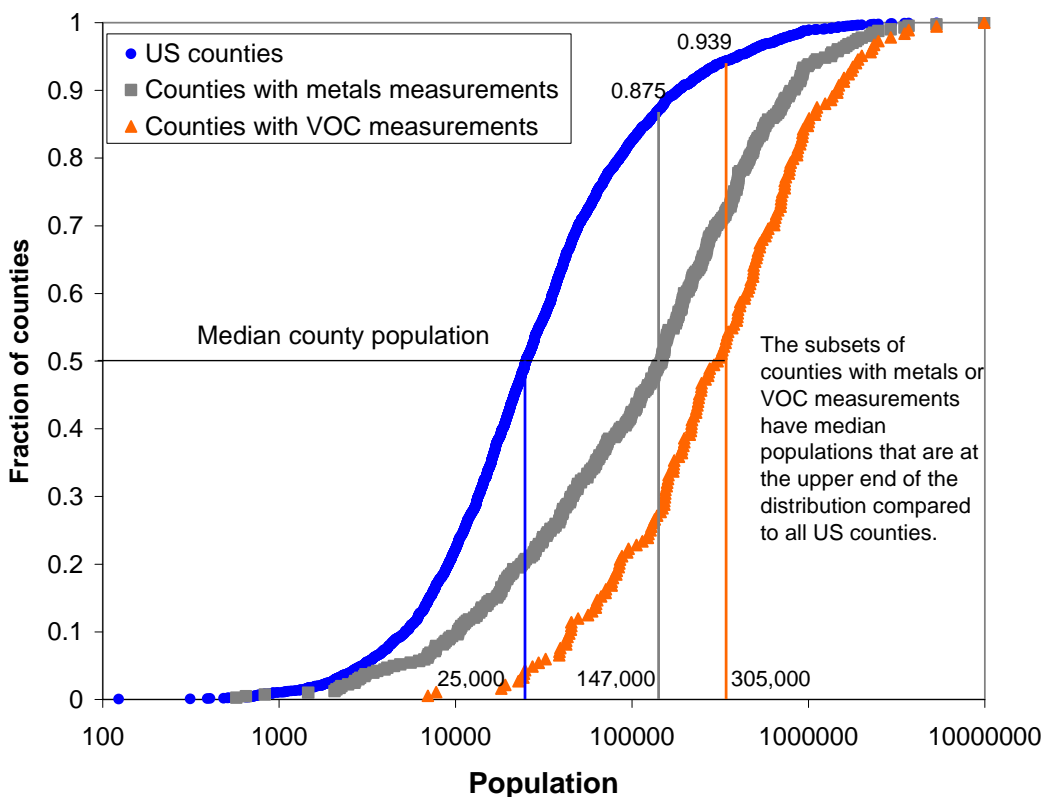

Figure 3. Cumulative distribution function illustrating the relative distribution of population of U.S. counties (blue circles) compared with the subsets of counties with metals measurements (gray squares) and VOC measurements (orange triangles) (2002 census estimates).

From counties with available measurements of ambient data, (1) small-urban estimates were developed using the 20th percentile concentrations; (2) medium-urban estimates were developed using the 50th percentile concentrations; and (3) large-urban estimates were generated using the 80th percentile concentrations. Counties lacking ambient data values were assigned to one of the four population bins based on their characteristics (i.e., population and urban/rural designation) and corresponding standard background concentration values were applied. Standard background concentration estimates are listed in **Table 4**.

Table 4. Standard estimates ($\mu g/m^3$) for use in NATA 2002 for counties of different populations.

Pollutant	Remote	Small-urban	Medium-urban	Large-urban
1,3-Butadiene	4.0×10^{-4}	2.6×10^{-2}	4.0×10^{-2}	6.6×10^{-2}
1,4-Dichlorobenzene	0.021	0.021	0.046	0.1
Acetaldehyde	0.16	0.57	0.90	1.2
Arsenic	1.3×10^{-4}	1.3×10^{-4}	3.5×10^{-4}	6.4×10^{-4}
Benzene	0.14	0.35	0.54	0.82
Chloroform	0.059	0.059	0.059	0.082
Chromium	4.1×10^{-5}	4.1×10^{-5}	2.9×10^{-4}	5.4×10^{-4}
Dichloromethane	0.11	0.11	0.19	0.30
Formaldehyde	0.20	0.69	1.2	1.7
Lead	4.9×10^{-4}	4.9×10^{-4}	1.0×10^{-3}	1.8×10^{-3}
Manganese	5.8×10^{-4}	5.8×10^{-4}	5.8×10^{-4}	1.0×10^{-3}
Nickel	6.5×10^{-5}	6.5×10^{-5}	1.5×10^{-4}	4.0×10^{-4}
Tetrachloroethylene	0.022	0.034	0.065	0.17
Toluene	0.041	0.45	0.87	1.5
Carbon tetrachloride	0.61	0.61	0.61	0.61
Methyl chloride	1.2	1.2	1.2	1.2
Methyl bromide	0.035	0.035	0.035	0.035
Methyl chloroform	0.17	0.17	0.17	0.17

Emissions-based Methodology

Sixteen of the 30 air toxics listed in Table 2 were identified as having inadequate ambient data to apply the ambient-based method. Background estimates based on the available ambient data for these pollutants would either have too few sites to extrapolate from or poor quality measurements on which to base background estimates. To provide spatially representative background concentrations for these pollutants, an alternative approach was required. The following emissions-based method was developed and is best applied to pollutants that are

emitted directly by a few large sources and that have short residence times in the atmosphere. The emissions-based method consists of four general steps:

1. Import emission inventory data into a geographic information system (GIS) and create emissions density maps.

2. Apply a spatial weighting scheme for deriving emissions gradients.

3. Normalize the emissions gradients.

4. Convert emissions gradient values to background concentration values.

The emissions-based method uses GIS technology to spatially weight and distribute county-level emissions estimates for each pollutant based on its residence time and air parcel transport potential. These county-level emissions gradient values were then post processed using lower- and upper-bound anchor points to convert emissions values to background concentrations.

Import Emission Inventory Data into a GIS and Create Emissions Density Maps

The 2002 county-level NEI data were imported into a GIS, and county-level emissions density maps were generated. Because the NEI data consist of a single emissions value for each county by pollutant, it is necessary to spatially distribute the emissions values across county boundaries to account for pollutant transport. To address this, emission inventory data were spatially weighted and distributed across county boundaries using a distance-residence time weighting scheme for each pollutant. To account for differences in pollutant lifetimes, or residence times (i.e., some pollutants remain in the air longer than others), a weighting function was derived and applied within the GIS to create emissions gradients for each pollutant. **Figure 4** shows the county-wide 2002 NEI data for ethylene dibromide.

Legend

2002 NEI Ethylene Dibromide

Tons/yr

- ● < 0.05
- ● 0.06 - 0.2
- ● 0.20 - 1
- ● 0.86 - 1.78
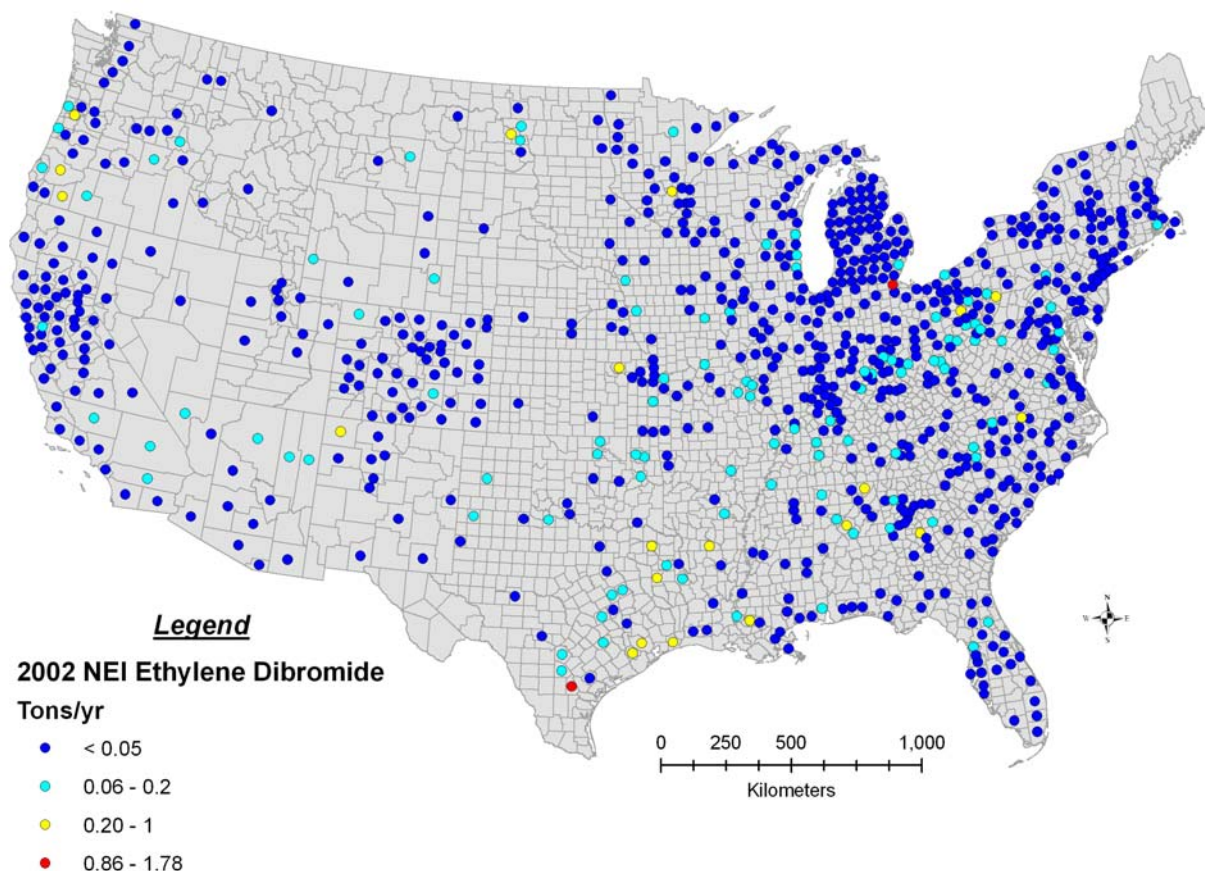
Figure 4. County-wide ethylene dibromide emissions (tons/year) as reported in the 2002 NEI. Each point on the map corresponds to the county's geographic center-point (centroid).

Development of Spatial Weighting Scheme for Deriving Emissions Gradients

The dispersion and dilution assumptions in a Gaussian plume dynamics model lead to concentration dilution of multiple orders of magnitude within a few kilometers. This approach is appropriate for modeling plume movement away from a discrete point source; however, it is less useful for modeling county-level transport. Based on an average wind speed of 3 m/s, air parcel transport is approximately 250 km per day. Consequently, significant transport can occur over two days for pollutants with long residence times. While pollution can be transported farther distances, it is likely that emissions contributions from counties at distances greater than 500 km will be relatively small.

Residence time is another contributing factor to pollutant concentrations over time. Chemical or physical removal competes with dilution if the residence time is on the same order of magnitude as the transport time. If pollutants are removed at rates much slower than they are diluted, they can be treated as inert on the time-scale of a few days. In contrast, if pollutants are removed on the time-scale of a few hours, then the removal processes will compete with dilution and the observed gradient in concentrations will be sharper.

For each pollutant of interest, buffer distances were calculated based on the residence time and dilution factors. For the dilution factor of a completely inert pollutant, a maximum buffer distance of 500 km was assumed to be the range of influence. While pollution can be transported around the globe, most point source emissions of pollution will have been fully diluted well within 500 km. This initial 500-km distance was then reduced as a function of the pollutant residence time. Equation 1 defines the drop-off as a function of distance,

$$B_x = \frac{500}{1.5^{\frac{0.5}{t}}}$$ (1)

where B_x is buffer distance and t is residence time in days. The exponential equation $1.5^{\wedge}(0.5/t)$ was empirically selected to provide buffer distances that reflect our expectations. **Table 5** summarizes the buffer distances computed using Equation 1.

Table 5. Calculated buffer distances for HAPs of interest.

Pollutants	Buffer Distance (km)
Hydrazine	222
Chromium (VI)	498
Ethylene Dichloride	498
Naphthalene	222
Propylene Dichloride	496
Ethylene Oxide	485
Acrylonitrile	482
Cadmium	490
Beryllium	490
Ethylene Dibromide	498
Benzidine	409
Quinoline	499
Bis(2-Ethylhexyl)Phthalate	500
1,2-Dibromo-3-Chloropropane	497
Trichloroethylene	483
1,1,2,2-Tetrachloroethane	499
Vinyl Chloride	451
Chloroprene	40
Acrolein	333
1,3-Dichloropropene	425

To provide a conceptual model of how the buffer distances are applied, consider chloroprene and cadmium. Chloroprene has a relatively short residence time and a resulting buffer distance of 40 km. Assume that the emissions point source for chloroprene is located at

the county centroid. As the distance away from the county centroid increases, the concentration of chloroprene will rapidly decrease due to dilution and chemical reaction. When the distance away from the county centroid equals 40 km, it is assumed that the concentration of chloroprene will equal zero. Therefore, the contribution of chloroprene from one county to another is likely to be small because this pollutant has a relatively short residence time. In contrast, cadmium has a much longer residence time and a buffer distance of 490 km. The concentration of cadmium does not reach zero until the distance away from the county centroid is 490 km; therefore, the contribution or influence of cadmium from one county to an adjacent one is likely to be relatively high.

For each pollutant, the buffer distance (B_X) was used in Equation 2 to estimate the fraction of emissions contribution from a particular county as the distance away from the county centroid increases.

$$f_1 = [(B_x - r)/B_x)]^2 \qquad \qquad \textbf{(2)}$$

where r is the distance between county centroids, B_x is the distance from the county centroid where the pollutant concentration equals zero, and f_1 is the fraction of emissions contribution from a specific county. The resultant value, f_1, is the fraction of the total emissions of a particular county that are transported to a nearby county.

Example Calculation

Contribution of chloroprene from County 1 assuming a distance away from the centroid of 30 km (r = 30km):
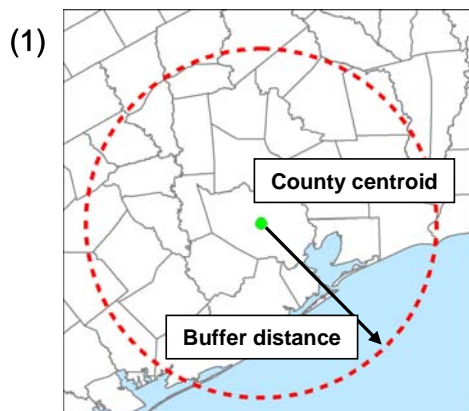
$f_1 = [(40km - 30km)/40km)]^2 = (0.25)^2 = 0.063$

Contribution of Cadmium from County 1 assuming a distance away from the centroid of 30 km (r = 30km):

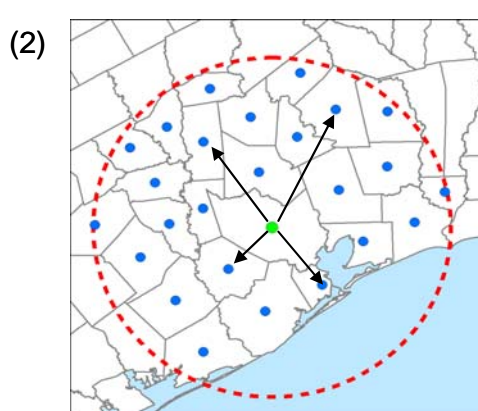$f_1 = [(490km - 30km)/490km)]^2 = (0.94)^2 = 0.882$

Figure 5 illustrates the process used to develop and apply the spatial weighting scheme.

(1) Calculate buffer distances for each pollutant and create buffers centered on county centroids.

(2) Calculate the distance(s) between all county centroids within the buffer. Each county centroid has an associated emissions value representing total county emissions for a specific pollutant.

(3) Use the following distance weighting equation to calculate the fraction of influence that each adjacent county has on the center county :

$$f_1 = [(Bx - r)/Bx\,)]^2$$

(4) Use the following summation equation to compute the additive influence of each emission value within the buffer:

$$F_{county} = \frac{\sum_{1}^{n} f_n E_n}{\max(\sum_{1}^{x} f_x E_x)}$$
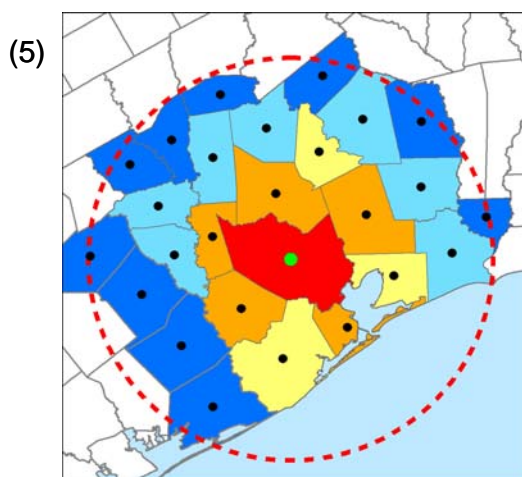
(5) Spatially distributed emissions values based on the spatial weighting scheme. Note that red indicates areas of high emissions influence and blue represents areas of low influence. Also note that emissions are higher in the center of the buffer zone because as the distance from the county centroid increases the pollutant emissions value decreases as a function of the pollutant buffer distance (listed in Table 5 by pollutant).

Figure 5. Illustration of the process used to apply the weighting scheme to spatially distribute county-level emissions.

Normalize the Emissions Gradients

Circular buffers centered on a county centroid were created within the GIS. The size of the buffer was determined by the buffer values corresponding to each pollutant listed in Table 5. The f_1 values for all counties were calculated within the GIS, and the combined contribution of each county was summed for a given buffer region. Equation 3 was then used to normalize the emissions contributions from all counties that influence a single county within the buffer zone.

15

$$F_{county} = \frac{\sum_{1}^{n} f_n E_n}{\max(\sum_{1}^{x} f_x E_x)}$$

(3)

where F_{county} is the county of interest, n is the number of counties with emissions that influence that county, f_n is the fraction emissions value calculated using Equation 2, E_n is the county emissions value from the 2002 NEI, x is the number of counties that influence the highest emissions county in the country, and max indicated the county with the highest emissions in the country for a given pollutant. This calculation is repeated for all counties with reported emissions by pollutant. The weighted emissions values for individual counties were summed and normalized using the county with the maximum emissions contribution (post-calculation). The resulting F_{county} is a unitless value between 0 and 1 representing the lowest and highest transport values in the country, respectively. The normalized F_{county} values were mapped to display the resulting emissions gradient by pollutant. The emissions gradient for each pollutant represents a unitless number corresponding to a range of emissions values. **Figure 6** shows an example map of a normalized emissions gradient field for ethylene dibromide.
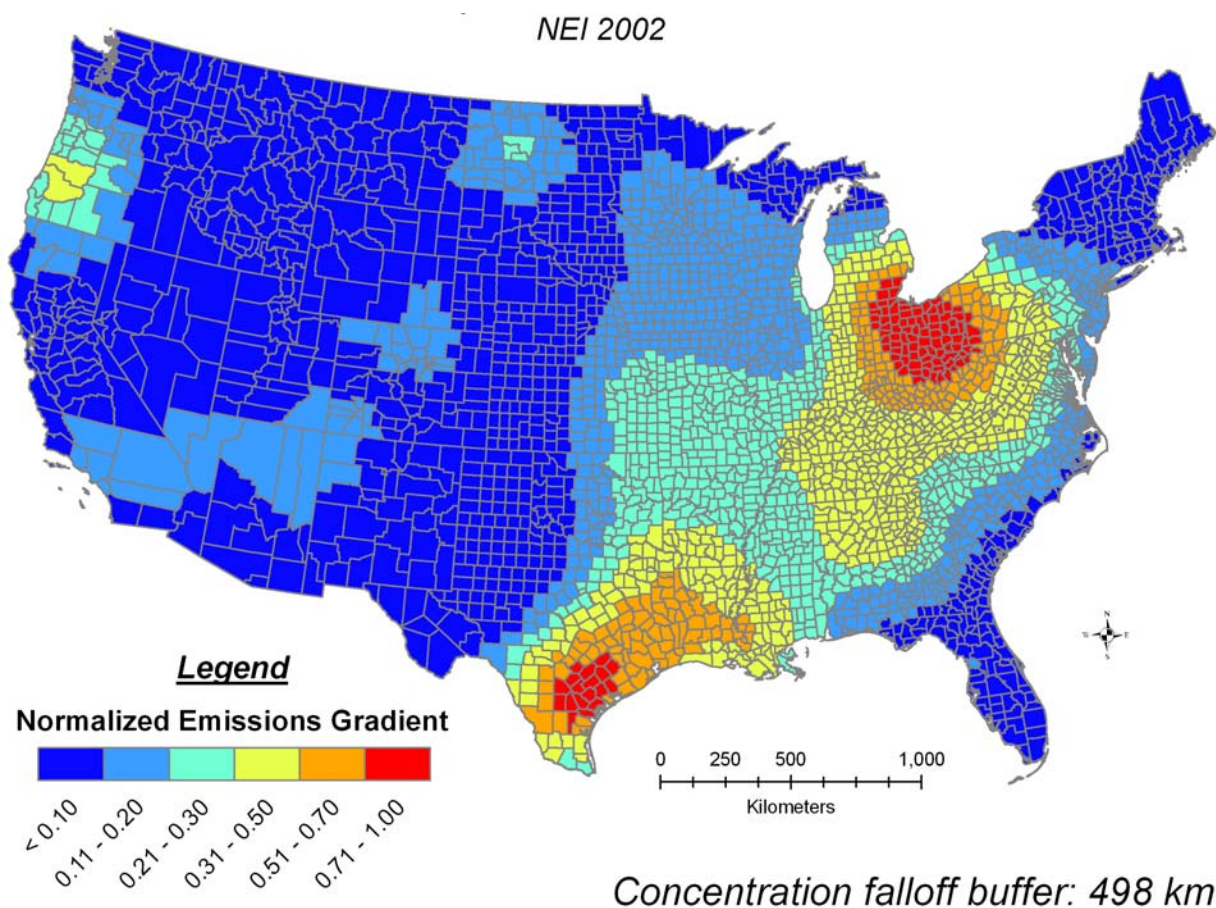
Figure 6. Example normalized emissions gradient map for ethylene dibromide.

16

Convert Emissions Gradient Values to Background Concentration Values

Emissions gradients were converted to the corresponding concentration values using minimum and maximum ambient concentration values. Estimating these minimum and maximum background concentration values requires multiple steps. The minimum ambient concentration represents a geographically remote concentration estimate. Remote concentrations for these pollutants are not typically measured or available in the published literature. Therefore, remote concentrations were estimated based on the 2002 NEI and a comparison to a pollutant with measured remote concentrations. Equation 4 shows the relationship used to derive these remote estimates.

$$[C_i] = \frac{E_i * t_i * [C_{tetrachloroethylene}]}{E_{tetrachloroethylene} * t_{tetrachloroethylene_i}} \tag{4}$$

where *[C]* is the remote concentration, *E* is the 2002 NEI value in tons per year, *t* is the residence time in years, and *i* is the pollutant of interest. **Table 6** provides the calculations and estimated concentrations for selected species.

Table 6. Calculation of remote concentration estimates using residence times and 2002 NEI emissions.

18

Name	Residence time (days)	2002 Emissions (tons per year)	Emissions x residence time (tons)	Fraction of tetrachloroethylene	Measured remote concentration ($\mu g/m^3$)	Estimated remote concentration ($\mu g/m^3$)
Benzyl Chloride	3	325	2.7	0.004		9.3E-05
Ethylene Dibromide	50	24	3.2	0.005		1.1E-04
Vinyl Chloride	2	1306	7.1	0.011		2.5E-04
Naphthalene	0.25	14729	10.1	0.02		3.5E-04
1,2-Dichloropropane	30	160	13.2	0.02		0.0005
Ethylene Oxide	7	695	13.3	0.02		0.0005
Acrylonitrile	5.6	1024	15.7	0.02		0.0005
Bromoform	540.0	22	32.6	0.05		0.0011
1,3-Dichloropropene	1.25	11518	39.4	0.06		0.0014
Acrolein[1]	0.5	29647	40.6	0.06		0.0014
Ethylene Dichloride	42	453	52.0	0.08		0.0018
1,1,2,2-Tetrachloroethane	91.3	296	74.0	0.12		0.0026
1,1,2-Trichloroethane	49	851	114.1	0.18		0.0040
Bis(2-ethylhexyl) phthalate	200.0	271	148.5	0.23		0.0052
Trichloroethylene	6	10808	177.5	0.28	0.005	0.0062
Carbon disulfide	7.0	15545	297.9	0.47		0.010
Xylenes	0.2	584519	320.1	0.51		0.011
1,4-Dichlorobenzene	31.0	7231	613.7	0.97		0.021
Tetrachloroethylene	6.5	35577	633.1	1.00	0.022	0.022
Toluene	0.5	891520	1220.4	1.93	0.041	0.042
Chloroform	80.0	6782	1485.5	2.35	0.059	0.052
Benzene	3.0	410892	3374.9	5.33	0.140	0.12
Dichloromethane	30.0	51057	4193.6	6.62	0.110	0.15
Carbon tetrachloride[2]	10950.0	454	13620.3	21.51	0.610	0.47
Methyl bromide	365.0	14777	14766.5	23.32	0.056	0.51

[1] Acrolein is formed secondarily and may not be well represented using primary emissions estimates.
[2] Carbon tetrachloride has a very long residence time, which makes predictions based on current emissions moot.

Maximum concentrations used to develop scaling factors are based on the ASPEN model-predicted pollutant concentration for the county with the highest concentration. We expect this county to contribute the most to background concentrations in adjacent counties. However, we also expect the concentrations from this county to be diluted during transport to the adjacent and downwind counties. Therefore, the 10th percentile concentration in the highest county in the United States from the ASPEN model predictions from NATA 1999 was used as the highest background concentration in the county.

The minimum and maximum background concentration estimates were used to develop a linearly interpolated scaling system to apply to the emissions gradient data. A simple linear relationship was derived relating the minimum and maximum concentrations to the highest and lowest emissions gradient values for each county. In a linear slope equation of $y = mx + b$, b is equal to the remote background estimate and m is equal to the maximum background estimate minus the minimum background estimate. This equation was used to predict (or convert) emissions gradient values to concentration values for all counties in the United States by setting x to the county-specific normalized emissions gradient value and solving the equation for y.

Figure 7 shows an example of how the normalized emissions gradients were associated with ambient concentrations using a linear interpolation approach for ethylene dibromide. In this figure, the two endpoint concentrations are anchored at the highest (1.0) and lowest (0.0) normalized emissions gradient for ethylene dibromide. All other county background concentrations are then interpolated between these two points based on their emissions gradient values.
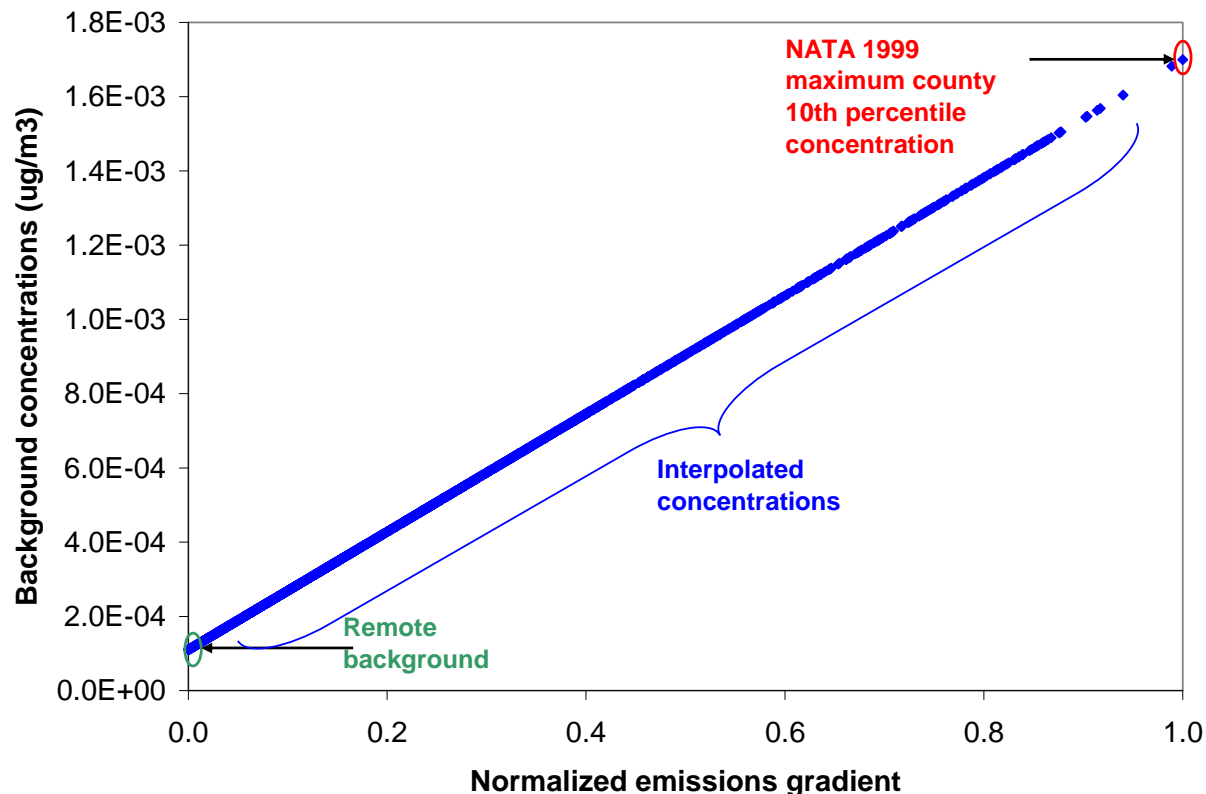
Figure 7. Example of how the normalized emissions gradient data were converted
to concentration data using a linear interpolation approach.

An automated GIS-based methodology was used to carry out the emissions-based method
to allow for rapid creation of these values. The GIS-based approach was scalable across
pollutants. The 16 HAPs selected for the emissions-based method were identified using the
criteria discussed in the Introduction *and* the following criteria:

- Pollutants were poorly characterized by ambient concentrations (i.e., lack of monitoring
  sites or MDLs insufficient to characterize concentrations).

- Pollutants are likely contributors to cancer risk or non-cancer hazard values based on
  previous NATA assessments (i.e., on the background list or risk driver/contributor list).

- Pollutants are predominantly emitted by large stationary sources:

  - Due to the nature of the methodology, the emissions-based method is most applicable
    to pollutants emitted by a few large sources because the emissions gradients were
    developed assuming that the emissions are released at a specific point and decrease as
    the distance from the point increases. Ubiquitous sources such as motor vehicles or
    area sources are unlikely to be represented well by this methodology because the
    gradients will be more diffuse.

  - Based on natural breaks in the NEI data distribution, most of the proposed pollutants
    were (1) emitted in fewer than 50 counties nationwide, (2) emitted in fewer than

150 counties that contributed two-thirds of total emissions, or (3) dominated by emissions sources other than point sources.

- Pollutants are not produced secondarily, biogenically, and/or have a residence time exceeding one year. The 2002 NEI will not be representative of the actual emissions sources and concentrations for these types of pollutants using this method because emission inventories only report primary emissions.

Table 7 lists candidate air toxics and the two criteria that were applied to determine candidate pollutants for the emissions-based method. Column 1 lists the candidate pollutants, column 2 lists the number of counties that contribute at least two-thirds of total emissions nationally (from point sources), and column 3 lists the known sources of these pollutants that would not be accounted for by the NEI. Pollutants emitted primarily by non-point source emissions (i.e., area and mobile sources) are less likely to be represented well by the emissions-based method. In particular, secondary production or pollutants with very long residence times are not considered good candidates for this method.

To identify pollutants that are likely to have maximum "background" concentrations that exceed 10^{-6} risk, an analysis of the NATA 1999 results was performed on the pollutants listed in Table 6 to assess the magnitude of concentrations expected in counties surrounding high emissions areas. This analysis was done by examining the 10th percentile concentration minus background contribution for the highest concentration counties. These concentrations were then compared to the relevant benchmark to illustrate the potential risk/hazard associated with these background concentrations. **Table 8** lists the estimated background contributions derived by this approach compared to the NATA 1999 background estimates.

The typical county will likely have far lower background concentrations than those estimated in Table 8. A list of prioritized HAPs for which to perform the emissions-based approach was created based on the summary of analyses in Table 6 and the following criteria:

1. The estimated risk/hazard in the highest counties and the risk/hazard associated with previous NATA results.

2. The suitability of emissions sources (i.e., point-sources and primary emissions) and the number of high-emitting counties.

21

Table 7. Candidate air toxics for the emissions-based approach.

Pollutants	Point Source dominated (# counties >66% of total)	Non NEI Sources?
Chloroprene	2	No
Chlordane	4	No
Hexachloroethane	6	Yes (decadal residence time)
Propylene Dichloride	7	No
1,2-Dibromo-3-Chloropropane	8	No
Hydrazine	10	No
Hexachlorobutadiene	10	No
Carbon disulfide	10	No
Benzidine	11	No
Phosgene	11	Yes (secondary)
Ethylene Dichloride	18	No
Quinoline	25	No
Bis(2-Ethylhexyl)Phthalate	36	No
Acrylonitrile	47	No
Beryllium	56	No
Vinyl Chloride	61	No
Ethylene Dibromide	74	No
Chromium (VI)	85	No
Benzyl Chloride	100	No
1,1,2,2-Tetrachloroethane	109	No
Cadmium	133	No
Trichloroethylene	149	No
Acrolein	<50% point source emissions	Yes (secondary production)
Naphthalene	<50% point source emissions	No
Ethylene Oxide	<50% point source emissions	No
1,1,2-Trichloroethane	<50% point source emissions	No
1,3-Dichloropropene	<50% point source emissions	No
Lindane	<50% point source emissions	No

Table 8. Maximum contribution to background risk for counties near emissions based on the highest county's 10th percentile NATA 1999 concentration. Pollutants in italics were not risk drivers or contributors in NATA 1999.

Pollutants	NATA 1999 Characterization	Est'd 2002 upper background risk (hazard)
Chromium (VI)	Regional risk driver	100.0
Acrolein	National noncancer driver	25.0
Hydrazine	Regional risk driver	10.0
Naphthalene	Regional risk driver	3.0
Ethylene Oxide	Regional risk driver	2.8
Cadmium	Regional noncancer driver	2.5
Trichloroethylene	Regional risk contributor	2.2
Acrylonitrile	Regional risk contributor	2.0
Beryllium	Regional risk contributor	2.0
1,1,2-Trichloroethane		1.5
1,3-Dichloropropene	Regional risk contributor	1.3
Ethylene Dibromide	National risk contributor	1.0
1,1,2,2-Tetrachloroethane	National risk contributor	0.5
1,2-Dibromo-3-Chloropropane		0.5
Ethylene Dichloride	National risk contributor	0.5
Quinoline	Regional risk contributor	0.5
Benzyl Chloride		0.4
Propylene Dichloride	Regional risk contributor	0.4
Vinyl Chloride	Regional risk contributor	0.3
Bis(2-Ethylhexyl)Phthalate	National risk contributor	0.1
Lindane		0.1
Hexachlorobutadiene		0.1
Hexachloroethane		0.1
Chloroprene		0.0
Benzidine	Regional risk driver	0.0
Carbon disulfide		0.0
Phosgene		0.0
Chlordane		0.0

Table 9 lists the final set of candidate air toxics for the emissions-based approach in ranked order. The pollutants are color-coded to highlight breakpoints that appear in the data.

Table 9. Candidate air toxics for which background concentrations were estimated using the emissions-based method. Pollutants are ordered using the ranking criteria described above.

Pollutants	NATA 1999 Characterization	Est'd 2002 upper background risk (hazard)	Rank
Hydrazine	Regional risk driver	10.0	4.5
Chromium (VI)	Regional risk driver	100.0	7
Ethylene Dichloride	National risk contributor	0.5	9.75
Naphthalene	Regional risk driver	3.0	9.75
Propylene Dichloride	Regional risk contributor	0.4	10
Ethylene Oxide	Regional risk driver	2.8	10
Acrylonitrile	Regional risk contributor	2.0	10.25
Cadmium	Regional noncancer driver	2.5	10.25
Beryllium	Regional risk contributor	2.0	10.75
Ethylene Dibromide	National risk contributor	1.0	10.75
Benzidine	Regional risk driver	0.0	10.75
Quinoline	Regional risk contributor	0.5	11.5
Bis(2-Ethylhexyl)Phthalate	National risk contributor	0.1	11.75
1,2-Dibromo-3-Chloropropane		0.5	12
Trichloroethylene	Regional risk contributor	2.2	12
1,1,2,2-Tetrachloroethane	National risk contributor	0.5	12.5
Vinyl Chloride	Regional risk contributor	0.3	13.5
Chloroprene		0.0	13.75
Acrolein	National noncancer driver	25.0	13.75
1,3-Dichloropropene	Regional risk contributor	1.3	14
Hexachlorobutadiene		0.1	14.75
Chlordane		0.0	15
Carbon disulfide		0.0	16
1,1,2-Trichloroethane		1.5	16.5
Benzyl Chloride		0.4	16.75
Lindane		0.1	19.25
Hexachloroethane		0.1	19.25
Phosgene		0.0	22

RESULTS

Background concentrations for all of the pollutants listed in Table 1 were estimated using either the ambient- or the emissions-based method. Results from the ambient-based method are described, followed by the results of the emissions-based method.

Ambient-based Method Results

The results of the ambient-based method provide county-level background concentration estimates for a few hundred counties in the United States and associated territories. For all other counties, standard background concentration estimates were assigned (as listed in Table 4) based on the characteristics of the county (i.e., remote, small-urban, medium-urban, and large-urban). The ambient-based results can be best examined by looking at the counties for which background concentrations will be assigned to one of the four bins. **Table 10** provides summary data on the number of counties and corresponding populations living in counties assigned to the four bins. Overall, about 80% of all counties fall into the remote or small-urban categories. However, as a function of population, the remote and small-urban counties account for only 5.3% and 22.2% of the U.S. population, respectively. In contrast, while the 35 largest urban counties account for 25% of the population, they only make up 1% of the total number of counties. The medium-urban counties account for the remaining 47.4% of the U.S. population.

Table 10. Statistics on population bins used for classifying background concentrations.

County Type	Number of Counties	Total Population	Average Population
Remote	1,354	15,526,420	11,467
Small urban	1,385	64,955,977	46,900
Medium urban	448	138,446,014	309,031
Large urban	35	73,078,709	2,087,963

The standard urban estimates listed in Table 4 are weighted by their unit risk estimates as provided by EPA OAQPS (U.S. Environmental Protection Agency, 2004) and displayed cumulatively as risk-weighted concentrations in **Figure 8**. The magnitude of each bar indicates the cumulative risk-weighted value for each pollutant in each of the standard estimates. Cumulative risk-weighted background concentrations for these pollutants increase from 11.6 per million in remote areas to 26.8 per million in large urban areas. In remote and small-urban areas, the cumulative risk-weighted concentrations are dominated by carbon tetrachloride, with a smaller contribution from benzene. In the medium- and large-urban areas, contributions from benzene, arsenic, acetaldehyde, chromium, and 1,3-butadiene are all above 1-in-a-million. The cumulative risk-weighted concentrations from these pollutants are higher than results from NATA 1999 and NATA 1996, largely as a result of the addition of the metals background concentrations. Metals were not included in background estimates for either of the two previous iterations and their predicted concentrations were among the most underestimated of all air toxics.
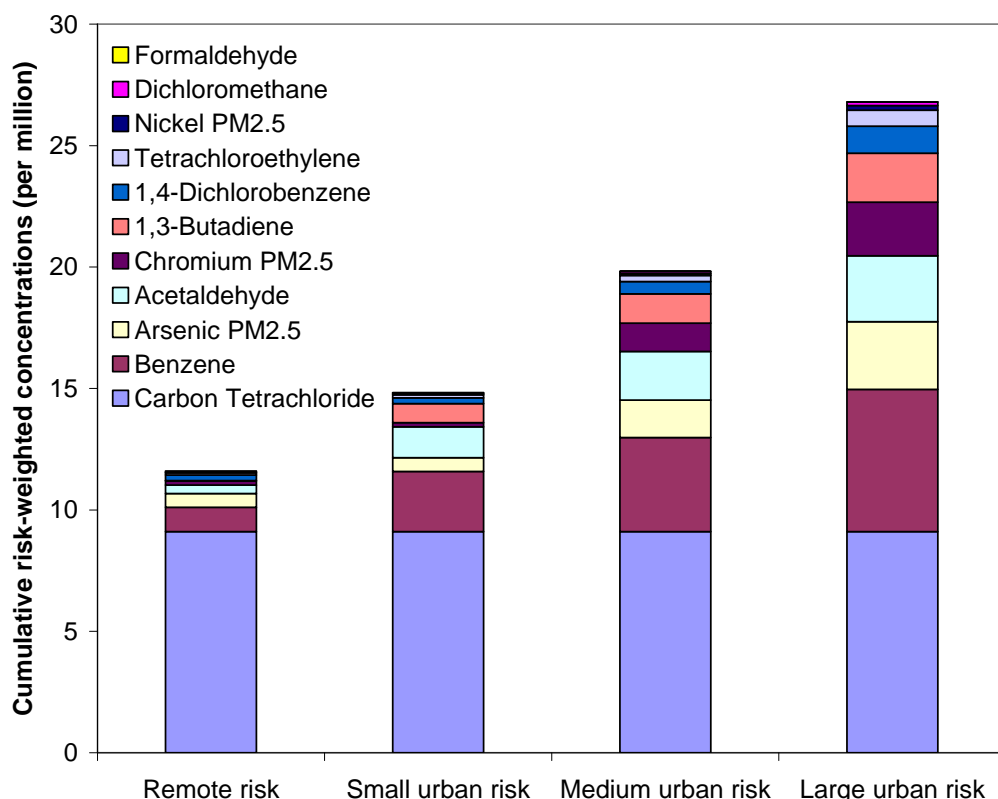
25

Figure 8. Cumulative risk-weighted concentrations of ambient-methods air toxics pollutants and carbon tetrachloride across the standard bins.

Figure 8 indicates that background concentrations will contribute more than 10-in-a-million risk to the total population, and more than 20-in-a-million risk to almost two-thirds of the U.S. population. The largest contributor to this risk is carbon tetrachloride which has been phased out under the Kyoto Protocol and is no longer produced in the United States. All other pollutants continue to be emitted throughout the United States, and most of their influence is a result of downwind transport.

Hazard-weighting of the air toxics was also performed and the results are shown in **Figure 9**; concentrations are weighted by chronic non-cancer reference concentrations. The cumulative hazard-weighted background concentrations are all below the hazard quotient level of one, which indicates that background concentrations for these pollutants are not expected to be at levels of non-cancer concern. In urban areas, the largest contributors to cumulative hazard among these pollutants are formaldehyde and acetaldehyde.
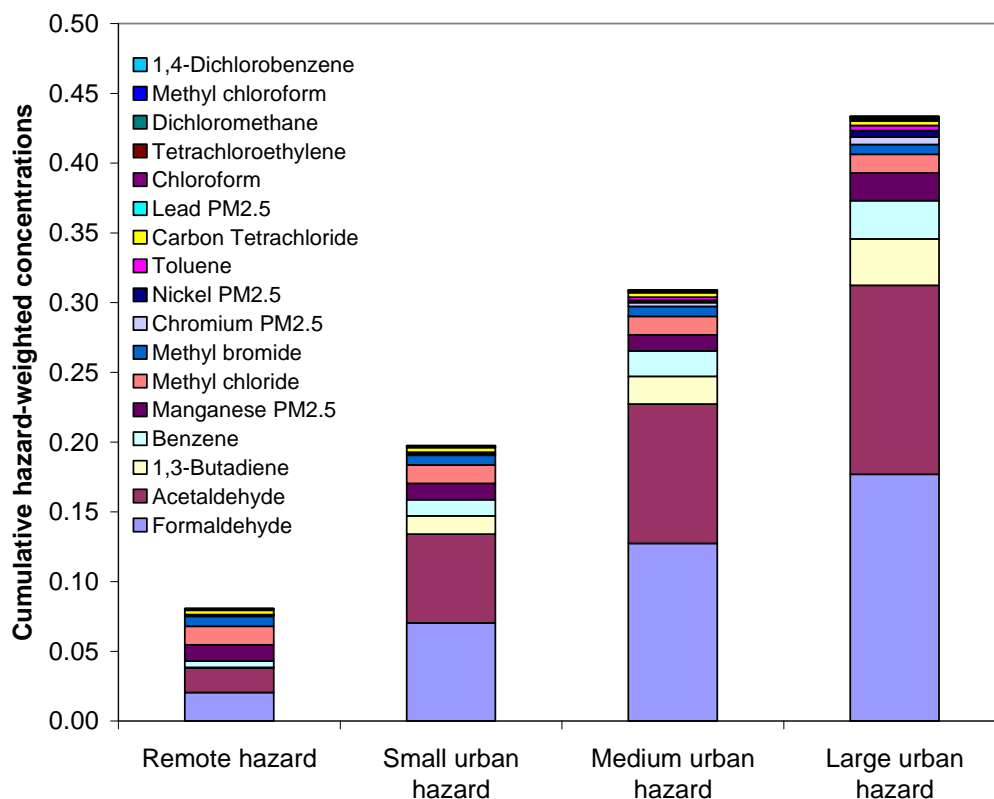
Figure 9. Cumulative hazard-weighted concentrations of ambient-methods air toxics pollutants across the standard-estimate bins and the spatially invariant pollutants.

Emissions-based Method Results

Normalized emissions gradients were created for 16 pollutants. These gradients were then scaled using the remote and maximum background estimates shown in **Table 11**. The results of the emissions-based method were also compared to chronic health benchmarks in Table 11. Pollutants with background concentrations that exceeded the health benchmarks in at least one county include acrylonitrile, beryllium, chromium VI, ethylene oxide, ethylene dibromide, and naphthalene. No pollutant concentrations exceeded the non-cancer reference concentration.

Table 11. Range of concentrations applied to emissions gradients. Shaded boxes indicated pollutants with concentrations that substantially exceeded the health benchmark in at least one county.

Pollutant	Remote estimate	1999 NATA max county 10th percentile	Cancer benchmark	RfC	CB remote	CB max	RFC max
Acrylonitrile	0.00055	0.0302	1.47E-02	2	0.0	2.1	0.02
Beryllium	1.80E-05	1.05E-03	4.10E-04	0.020	0.0	2.6	0.05
Benzidine	9.90E-09	6.78E-06	1.49E-05	10	0.0	0.5	0.00
Bis(2-ethylhexyl)phthalate	0.0052	7.77E-03	4.17E-01	10	0.0	0.0	0.00
Cadmium	3.70E-05	1.29E-04	5.50E-04	0.020	0.1	0.2	0.01
Chromium VI	2.20E-05	4.27E-03	8.33E-05	0.100	0.3	51.2	0.04
Dibromochloropropane	1.30E-06	9.05E-05	5.00E-04	0.200	0.0	0.2	0.00
Ethylene dibromide	1.10E-04	1.70E-03	1.66E-03	9	0.1	1.0	0.00
Ethylene dichloride	0.0018	1.98E-02	3.85E-02	2400	0.0	0.5	0.00
Ethylene oxide	4.63E-04	3.00E-02	1.14E-02	30	0.0	2.6	0.00
Hydrazine	1.30E-07	1.79E-04	2.04E-04	0.20	0.0	0.9	0.00
Naphthalene	3.50E-04	3.38E-01	2.94E-02	3	0.0	11.5	0.11
1,2-Dichloropropane	4.60E-04	1.05E-02	5.26E-02	4	0.0	0.2	0.00
Quinoline	8.80E-07	1.89E-03					
1,1,2,2-Tetrachloroethane	2.60E-03	0.00875	1.72E-02		0.2	0.5	
Trichloroethylene	0.005	1.75E-01	5.00E-01	600	0.0	0.3	0.00

CB = cancer benchmark, OAQPS
RFC = reference concentration (chronic non-cancer benchmark)

Figure 10 shows the number of counties for those pollutants that exceeded the cancer benchmark using the emissions-based method. The results indicate that a large number of counties exceed the cancer benchmark for chromium VI. This is inconsistent with measurements of chromium $PM_{2.5}$ for which concentrations are typically far lower than those estimated using this method. Given the discrepancy between the two methods, we think the ambient-based method produces more reliable values for chromium VI than does the emissions-based method. We recommend using the ambient-based method estimates of chromium VI at this time. For other air toxics, naphthalene was the only pollutant with a single high background prediction. However, the total number of counties exceeding the cancer benchmark is small. This is consistent with what would be expected given naphthalene transport times. It is possible that these discrepancies are the result of a single high prediction for the 10th percentile background concentrations in NATA 1999. These results should be compared to NATA 2002 before they are used.

Aside from the pollutants shown in Figure 10, the emissions-based method indicates that most of the pollutants investigated will not contribute substantially to health risk. Of those that do contribute, most are limited to large impacts in less than 3% of all U.S. counties. Only chromium VI is contributing in most counties, and it is likely that its impact is overestimated by

using the emissions-based method. These results make sense given the small quantities of these pollutants that are emitted relative to those that are more routinely measured.
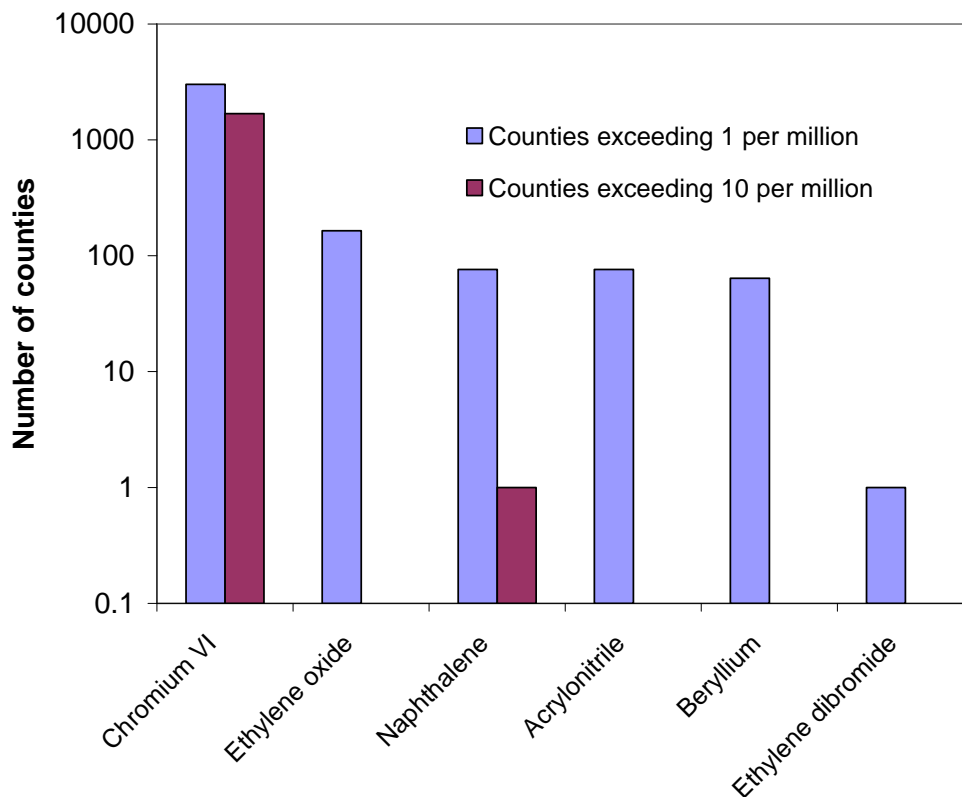

Figure 10. Number of counties exceeding the 1-in-a-million and 10-in-a-million cancer benchmark for emissions-based pollutants.

SUMMARY

Background concentrations were estimated for the NATA 2002 analysis. These estimates were developed using two methods: an ambient-based method and an emissions-based method. The method chosen for each pollutant was determined based on the availability and reliability of ambient data. Fourteen pollutant concentrations were estimated using the ambient-based method and 16 pollutant concentrations were estimated using the emissions-based method.

Overall, the ambient-based results indicate that carbon tetrachloride, benzene, arsenic, acetaldehyde, and chromium have background concentrations above the 10^{-6} cancer benchmark for large portions of the population. The emissions-based method indicates that chromium VI is above the health benchmark for most counties, an observation that conflicts with the results of the ambient-based method. Finally, the emissions-based method shows that ethylene oxide, naphthalene, acrylonitrile, and beryllium background concentrations above the 10^{-6} cancer benchmark in 50 to 150 counties.

REFERENCES

McCarthy M.C., Hafner H.R., and Montzka S.A. (2006) Background concentrations of 18 air toxics for North America. *J. Air and Waste Manag. Assoc.* **56**, 3-11 (STI-903550-2589).

Montzka S.A., Butler J.H., Elkins J.W., Thompson T.M., Clarke A.D., and Lock L.T. (1999) Present and future trends in the atmospheric burden of ozone-depleting halogens. *Nature* **398**, 690-694.

Montzka S.A., Spivakovsky C.M., Butler J.H., Elkins J.W., Lock L.T., and Mondeel D.J. (2000) New observational constraints for atmospheric hydroxyl on global and hemispheric scales. *Science* **288**, 500-503.

Prinn R.G., Weiss R.F., Fraser P.J., Simmonds P.G., Cunnold D.M., Alyea F.N., O'Doherty S., Salameh P., Miller B.R., Huang J., Wang R.H.J., Hartley D.E., Harth C., Steele L.P., Sturrock G., Midgley P.M., and McCulloch A. (2000) A history of chemically and radiatively important gases in air deduced from ALE/GAGE/AGAGE. *J. Geophys. Res* **105**, 17,751-717,792.

U.S. Environmental Protection Agency (2004) Technology transfer network air toxics web site - prioritized chronic dose-response values. Available on the Internet at <http://www.epa.gov/ttn/atw/toxsource/summary.html>.